ELSEVIER

Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE DIRECT®

Journal of Computational and Applied Mathematics 157 (2003) 365–382

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

Multigrid methods for parabolic distributed optimal control problems

Alfio Borzì

Institut für Mathematik, Karl-Franzens-Universität Graz, Heinrichstr. 36, A-8010 Graz, Austria

Received 12 March 2002; received in revised form 15 January 2003

Abstract

Multigrid schemes that solve parabolic distributed optimality systems discretized by finite differences are investigated. Accuracy properties of finite difference approximation are discussed and validated. Two multigrid methods are considered which are based on a robust relaxation technique and use two different coarsening strategies: semicoarsening and standard coarsening. The resulting multigrid algorithms show robustness with respect to changes of the value of v, the weight of the cost of the control, is sufficiently small. Fourier mode analysis is used to investigate the dependence of the linear twogrid convergence factor on v and on the discretization parameters. Results of numerical experiments are reported that demonstrate sharpness of Fourier analysis estimates. A multigrid algorithm that solves optimal control problems with box constraints on the control is considered.

MSC: 49K20; 65M06; 65M12; 65M55

Keywords: Optimal control problems; Parabolic partial differential equations; Finite differences; Multigrid methods

1. Introduction

A standard approach to the solution of parabolic problems is by time-stepping methods. These methods require to treat time evolution in a strict sequential way which may be computationally disadvantageous. In fact, when solving optimal control problems [15,16,18] which are governed by parabolic partial differential equations with opposite orientation, the sequential approach makes it difficult to implement the time coupling in the optimality system. Especially for this reason we consider parabolic multigrid methods that solve distributed parabolic optimal control systems in one

E-mail address: alfio.borzi@uni-graz.at (A. Borzì).

0377-0427/03/$ - see front matter
doi:10.1016/S0377-0427(03)00417-5

shot in the whole space–time cylinder. Multigrid schemes have been considered previously in, e.g., [7,10,11,13,20–22] to solve parabolic evolution problems without control.

The disadvantage of the space–time approach is the requirement to store the dependent variables for all time steps. This is certainly a limitation which arises when open-loop optimal control problems on large time intervals are considered. However, in most cases relatively small time intervals are of interest and usually memory is not an issue in modern computer architectures. Notice that the present approach may be used in the framework of receding horizon techniques [14] to solve optimal control problems in unbounded time intervals.

The main advantage of using the space–time multigrid strategy is the fact that these methods provide fast solvers whose convergence factors are mesh independent. Furthermore we show that the convergence behavior of our multi-grid schemes does not deteriorate as the weight of the cost of the control tends to be small.

Detailed discussions on space–time multigrid methods for parabolic problems may be found in [10,13,20,22]. Our main contribution to this field is the extension of these methods to the case of parabolic optimality systems which consist of two coupled parabolic equations with opposite orientation and a control equation. It results that a similar multigrid structure as used in the single equation case can be used in the present context except the smoothing iteration. In fact, for the problems considered here the design of a suitable smoothing iteration is not straightforward. One should guarantee a robust coupling between state and control variables and take care of the fact that the state variable evolves forward in time while the adjoint variable evolves backwards. We present a pointwise smoothing scheme that satisfies these requirements.

We combine our smoothing iteration together with two different coarsening strategies to construct multigrid methods for parabolic optimal control problems discretized by finite differences and backward Euler scheme. The coarsening strategies considered here are standard time coarsening and semicoarsening in space. As may be expected from results in [10,13] the semicoarsening strategy provides the most efficient multigrid scheme. Nevertheless, we demonstrate that the standard coarsening approach provides a competitive multigrid algorithm as the weight of the cost of the control becomes sufficiently small and in case sufficiently small time intervals are considered. The latter result is also known in case of space–time multigrid scheme for parabolic problems with one orientation [13]. The present investigation is limited to fixed coarsening strategies, however the pointwise smoother introduced here could be used in combination with the adaptive coarsening method presented in [13].

The present paper is devoted to distributed optimal control problems governed by linear parabolic equations. Notice that the algorithms discussed in this paper have been successfully applied [2,4,5] to solve singular (nonlinear) parabolic optimal control problems [16]. Comparing the numerical results of this paper with those of [2,4,5] we notice that the convergence behavior of our algorithms remains qualitatively similar also in the presence of strong nonlinearities. In this paper we discuss the approximation properties of finite difference discretization and analyze the convergence properties of our multigrid algorithms.

In the following section the problem of distributed optimal control for a parabolic equation is formulated and its approximation by finite differences and backward Euler scheme is discussed. In this framework we show $\mathcal{O}(h^2 + \delta t)$ accuracy of the numerical approximation to the optimal control solution. The multigrid techniques considered in this paper, with emphasis on the smoothing iteration, are described in Section 3. Section 4 is dedicated to numerical experiments. In Section 5 we use Fourier mode analysis to investigate the dependence of the computational performance of multigrid

on the weight of the cost of the control and on the discretization parameters. The results of this analysis confirm the observations made with numerical experiments. In Section 6, a parabolic optimal control problem with box constraints on the control is considered: We show the advantage of using W-cycle multigrid iteration. A section of conclusions completes these paper.

2. Parabolic distributed optimal control problems and their approximation

The purpose of this work is to present multigrid solvers for the following optimal control problem

$$\begin{cases} \min_{u\in L^2(Q)} J(y,u), \\ -\partial_t y + \Delta y = u & \text{in } Q=\Omega\times(0,T), \\ y(\mathbf{x},0)=y_0(\mathbf{x}) & \text{in } \Omega \text{ at } t=0, \\ y(\mathbf{x},t)=0 & \text{on } \Sigma=\partial\Omega\times(0,T), \end{cases} \tag{1}$$

where we take $y_0(\mathbf{x})\in H_0^1(\Omega)$. We consider cost functionals of the tracking type given by

$$J(y,u)=\frac{1}{2}\|y-z\|^2_{L^2(Q)}+\frac{v}{2}\|u\|^2_{L^2(Q)}. \tag{2}$$

Here, $v>0$ is the weight of the cost of the control and $z\in L^2(Q)$ denotes the desired state. Then there exists a unique solution $(y^*,u^*)=(y^*(u^*),u^*)$ to the optimal control problem above; see [15]. Corresponding to our setting we have $y^*(u^*)\in H^{2,1}(Q)$ where $H^{2,1}(Q)=L^2(0,T;H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(0,T;L^2(\Omega))$.

The solution to (1) is characterized by the following optimality system

$$-\partial_t y+\Delta y=u, \tag{3}$$

$$\partial_t p+\Delta p+(y-z)=0, \tag{4}$$

$$vu-p=0 \tag{5}$$

with initial condition $y(\mathbf{x},0)=y_0(\mathbf{x})$ for the state equation (evolving forward in time) and terminal condition $p(\mathbf{x},T)=0$ for the adjoint equation (evolving backward in time). Here, for convenience we dropped the $*$-notation. From (4) and (5) we have $p,u\in H^{2,1}(Q)$. The a priori knowledge of the regularity of solution is essential for the numerical analysis which follows. The fact that u attains the same regularity of p is due to the special form of (5), the optimality condition. We focus on cases where the optimality condition provides a scalar relation between u and p. This includes the case of constrained control; see Section 6.

In the following, the numerical solution of the optimality system (3)–(5) in the framework of finite differences and backward Euler scheme is considered. Define a sequence of grids $\{\Omega_h\}_{h>0}$ given by

$$\Omega_h=\{\mathbf{x}\in\mathbf{R}^2 : x_j=jh,\ j\in\mathbb{Z}\}\cap\Omega.$$

We assume that Ω is a square and that the values of the spatial mesh size h are chosen such that the boundaries of Ω coincide with grid lines. We call $\bar{\Omega}_h$ the mesh, Ω_h is the set of interior mesh-points, and Γ_h is the set of boundary mesh-points. The negative Laplacian with homogeneous Dirichlet

boundary conditions is approximated by the common five-point stencil as in [12] and denoted by $-\Delta_h$. For grid functions v_h and w_h defined on Ω_h we introduce the discrete $L^2(\Omega)$-scalar product

$$(v_h, w_h)_{L_h^2(\Omega_h)} = h^2 \sum_{\mathbf{x}\in\Omega_h} v_h(\mathbf{x}) w_h(\mathbf{x})$$

with associated norm $|v_h|_0 = (v_h, v_h)^{1/2}_{L_h^2(\Omega_h)}$.

We require as well the discrete H^1-product given by

$$|v_h|_1 = \left(|v_h|_0^2 + \sum_{i=1}^{2} |\partial_i^- v_h|_0^2 \right)^{1/2},$$

where ∂_i^- denotes the backward difference quotient in the x_i direction and v_h is extended by 0 on grid points outside of Ω. The spaces L_h^2 and H_h^1 consist of the sets of grid functions v_h endowed with $|v_h|_0$, respectively $|v_h|_1$, as norm. For the definition of H_h^2 we refer to [12], as well. We have the inverse property $|y_h|_2 \leqslant ch^{-1}|y_h|_1$.

Functions in $L^2(\Omega)$ and $H^2(\Omega)$ are approximated by grid functions defined through their mean values with respect to elementary cells $[x_1 - h/2, x_1 + h/2] \times [x_2 - h/2, x_2 + h/2]$. This gives rise to the restriction operators $\tilde{R}_h : L^2(\Omega) \to L_h^2$ and $R_h : H^2(\Omega) \cap H_0^1(\Omega) \to L_h^2$ defined in [12]. The following property can be proved

$$|\tilde{R}_h v - R_h v|_0 \leqslant ch^2 |v|_{H^2(\Omega)} \quad \text{for all } v \in H^2(\Omega). \tag{6}$$

Here and below, c denotes a positive constant which does not depend on the discretization parameters.

Let $\delta t = T/N_t$ be the time step size. Define

$$Q_{h,\delta t} = \{(\mathbf{x}, t_m) : \mathbf{x} \in \Omega_h,\ t_m = (m-1)\delta t,\ 1 \leqslant m \leqslant N_t + 1\}.$$

On this grid, y_h^m denotes a grid function at time level m. The action of the time difference operator on this function is denoted by

$$\partial_t^+ y_h^m = \frac{y_h^m - y_h^{m-1}}{\delta t} \quad \text{and} \quad \partial_t^- y_h^m = \frac{y_h^{m+1} - y_h^m}{\delta t}.$$

For grid functions defined on $Q_{h,\delta t}$ we use the discrete $L^2(Q)$ scalar product with norm $\|v_{h,\delta t}\|_0 = (v_{h,\delta t}, v_{h,\delta t})^{1/2}_{L^2_{h,\delta t}(Q_{h,\delta t})}$.

For convenience, it is assumed that there exist positive constants $c_1 \leqslant c_2$ such that $c_1 h^2 \leqslant \delta t \leqslant c_2 h^2$. Hence h can be considered as the only discretization parameter. Therefore, in the following, the subscript δt is omitted.

On the cylinder Q_h define the family of functions piecewise constant on intervals $[t_m, t_{m+1})$ as follows:

$$V_h = \{v_h \mid v_h(t) = v_h(t_m) \text{ for } t \in [t_m, t_{m+1}),\ v_h(t_m) \in L_h^2(\Omega_h)\}.$$

The space–time extension of the operators $\tilde{R}_h$ and R_h are denoted by

$$\tilde{R}_{h,Q} : L^2(Q) \to V_h \quad \text{and} \quad R_{h,Q} : H^{2,1}(Q) \to V_h.$$

Condition (6) implies

$$\|\tilde{R}_{h,Q} v - R_{h,Q} v\|_0 \leqslant ch^2 |v|_{H^{2,1}(Q)}. \tag{7}$$

The discrete optimal control problem is specified next

$$\min \tfrac{1}{2}\,\|y_h - \tilde{R}_{h,Q} z\|_0^2 + \tfrac{\nu}{2}\,\|u_h\|_0^2, \quad -\partial_t^+ y_h + \Delta_h y_h = u_h. \tag{8}$$

Let $u_h^* \in L_h^2(Q_h)$ denote the unique solution to (8) and set $y_h^* = y_h(u_h^*)$. The optimality system related to (8) is found to be

$$\begin{aligned} &-\partial_t^+ y_h^* + \Delta_h y_h^* = u_h^*, \\ &\partial_t^- p_h^* + \Delta_h p_h^* = -(y_h^* - \tilde{R}_{h,Q} z), \\ &\nu u_h^* - p_h^* = 0. \end{aligned} \tag{9}$$

Next we eliminate u_h^* from this system and drop the superscript $*$. In expanded form, we obtain

$$\begin{aligned} &-[1+4\gamma] y_{ijm} + \gamma[y_{i+1jm} + y_{i-1jm} + y_{ij+1m} + y_{ij-1m}] + y_{ijm-1} \\ &\quad = \frac{\delta t}{\nu}\, p_{ijm}, \quad 2 \leqslant m \leqslant N_t + 1, \end{aligned} \tag{10}$$

$$\begin{aligned} &-[1+4\gamma] p_{ijm} + \gamma[p_{i+1jm} + p_{i-1jm} + p_{ij+1m} + p_{ij-1m}] + p_{ijm+1} \\ &\quad + \delta t(y_{ijm} - \tilde{z}_{ijm}) = 0, \quad 1 \leqslant m \leqslant N_t, \end{aligned} \tag{11}$$

where $\gamma = \delta t/h^2$, $2 \leqslant i, j \leqslant N_{\mathbf{x}}$ index the internal grid points and $\tilde{z} = \tilde{R}_{h,Q} z$. The implementation of the boundary conditions on Σ, of the initial condition at $t = 0$, and of the terminal condition at $t = T$ should be clear.

Now we can use the theory of Malanowski [17] to prove that the solution of (9) is second-order accurate. For this purpose we need to extend the approach of Malanowski [17] to the present finite difference framework. We now outline the proof and refer to [17] for all the details.

Using Lemma 1.1 and the approach of Theorem 1.2 of Malanowski [17], and (6) we have

$$\|u_h^* - R_{h,Q} u^*\|_0 \leqslant ch^2. \tag{12}$$

Next, assume that y be solution of the state equation with any given $u \in L^2(Q)$ and let y_h be its finite difference approximation where $u_h = \tilde{R}_{h,Q} u$. Then the following estimates hold

$$\|y_h - R_{h,Q} y\|_0 \leqslant ch^2 |y|_{H^{2,1}(Q)}$$

and in a similar way one has that $\|p_h - R_{h,Q} p\|_0 \leqslant ch^2 |p|_{H^{2,1}(Q)}$. Using these estimates and (12), we obtain the following estimates for the approximation of the state variable and of the adjoint variable

$$\|y_h^* - R_{h,Q} y^*\|_0 \leqslant ch^2 \quad \text{and} \quad \|p_h^* - R_{h,Q} p^*\|_0 \leqslant ch^2. \tag{13}$$

To validate the accuracy estimates (12) and (13), consider the following exact solution

$$y(\mathbf{x}, t) = t^2(1-t)^2 \sin(\pi x) \sin(\pi y),$$

$$p(\mathbf{x}, t) = 2\nu(1-t)t(\pi^2 t^2 - (\pi^2 - 2)t - 1) \sin(\pi x) \sin(\pi y)$$

for the optimal control problem with objective function given by

$$z(\mathbf{x}, t) = ((t-1)^2 t^2 + 2\nu(2\pi^4 t^4 - 4\pi^4 t^3 + 2(\pi^4 - 3)t^2 + 6t - 1)) \sin(\pi x) \sin(\pi y),$$

in $\Omega = (0,1) \times (0,1)$ and $T = 1$.

Table 1
Accuracy results: $s=1$, $v=10^{-4}$, $\delta t=32h^2$

$N_x \times N_y \times N_t$	$\|y-y_h\|_0$	$\|p-p_h\|_0$
$32\times 32\times 32$	$2.63\cdot 10^{-5}$	$5.64\cdot 10^{-7}$
$64\times 64\times 128$	$7.05\cdot 10^{-6}$	$1.47\cdot 10^{-7}$
$128\times 128\times 512$	$1.78\cdot 10^{-6}$	$3.73\cdot 10^{-8}$

In Table 1 results of numerical experiments with this choice of $z(\mathbf{x},t)$ and using $y_0(\mathbf{x})=y(\mathbf{x},0)$ are reported. We observe second-order convergence. In fact, the solutions errors reduce approximately as a factor of four by halving the space mesh size.

3. Multigrid schemes for parabolic optimal control problems

To solve parabolic optimal control problems we consider two multigrid schemes corresponding to two different coarsening strategies. These methods solve (10) and (11) for all time levels simultaneously. For this purpose, consider L grid levels where each level is indexed by $k=1,\dots,L$. We use a set of grids with space mesh size $h=h_k=h_1/2^{k-1}$, thus we have standard coarsening in the space directions. In the time direction we set $\delta t=\delta t_k=\delta t_1/s^{k-1}$, $s\in\{1,2\}$. If $s=1$ we have semicoarsening in space; the case $s=2$ is referred to as standard time coarsening. The choice of different coarsening strategies can be motivated by memory needs. Clearly, larger values of the coarsening factor s results in less memory requirements. Memory complexity for the two choices of s is proportional to the following

$$\sum_{k=1}^{L}\left(\frac{1}{2}\right)^{2(L-k)} \quad (s=1), \qquad \sum_{k=1}^{L}\left(\frac{1}{2}\right)^{3(L-k)} \quad (s=2).$$

The mesh of level k is denoted by $Q_k=Q_{h_k,\delta t_k}$. Any operator and variable defined on Q_k is indexed by k. Denote by $\mathbf{M}_k=M_k\times\tilde{M}_k$ the space of the system of nodal functions (v_k,w_k) which are zero on the lateral boundary and satisfy the initial and terminal condition.

The algebraic problem given by (10) and (11) at level L with given initial, terminal, and boundary conditions is represented by the following equation:

$$\mathbf{A}_L(\boldsymbol{\phi}_L)=\mathbf{f}_L. \tag{14}$$

where $\boldsymbol{\phi}_L$ is the pair $(y_L,p_L)\in\mathbf{M}_L$. Let $\mathbf{S}_k:\mathbf{M}_k\to\mathbf{M}_k$, $k=1,\dots,L$, be a smoothing operator. Then the multigrid FAS algorithm (see [6]), expressed in terms of the iteration operator B_L, to solve (14) in recursive form is given as follows.

FAS multigrid cycle

Set $\mathbf{B}_1\approx\mathbf{A}_1^{-1}$ (e.g., iterating with $\mathbf{S}_1$). For $k=2,\dots,L$ define $\mathbf{B}_k:\mathbf{M}_k\to\mathbf{M}_k$ in terms of $\mathbf{B}_{k-1}$ as follows. Let $\mathbf{g}\in\mathbf{M}_k$ and $\mathbf{q}^0=0$.

1. Set $\boldsymbol{\phi}^{(0)}=\tilde{\boldsymbol{\phi}}$ (starting approximation).
2. Define $\boldsymbol{\phi}^{(l)}$ for $l=1,\dots,m_1$, by

$$\boldsymbol{\phi}^{(l)}=\mathbf{S}_k(\boldsymbol{\phi}^{(l-1)},\mathbf{g}).$$

3. Set $\boldsymbol{\phi}^{\text{new}} = \boldsymbol{\phi}^{(m_1)} + \mathbf{I}_{k-1}^{k}(\mathbf{q}^m - \mathbf{I}_k^{k-1}\boldsymbol{\phi}^{(m_1)})$ where $\mathbf{q}^i$ for $i = 1,\dots,n$ is defined by

$$\mathbf{q}^i = \mathbf{q}^{i-1} + \mathbf{B}_{k-1}[\mathbf{I}_k^{k-1}(\mathbf{g} - \mathbf{A}_k(\boldsymbol{\phi}^{(m_1)})) + \mathbf{A}_{k-1}(\hat{\mathbf{I}}_k^{k-1}\boldsymbol{\phi}^{(m_1)}) - \mathbf{A}_{k-1}\mathbf{q}^{i-1}].$$

4. Set $\mathbf{B}_k\mathbf{g}=\boldsymbol{\phi}^{(m_1+m_2+1)}$ where $\boldsymbol{\phi}^{\ell}$ for $\ell=m_1+2,\dots,m_1+m_2+1$, is given by Step 2 with $\boldsymbol{\phi}^{(m_1+1)}=\boldsymbol{\phi}^{\text{new}}$.

Notice that we can perform n two-grid iteration at each working level. For $n = 1$ we have a $V(m_1, m_2)$-cycle and for $n = 2$ we have a $W(m_1, m_2)$-cycle; n is called the cycle index [19]. In the following, N is the number of FAS V- or W-cycles that are applied to solve the problem at hand.

We choose $\mathbf{I}_k^{k-1}$ to be the half-weighted restriction operator in space with no averaging in the time direction. It is given in stencil form by

$$\mathbf{I}_k^{k-1} = \frac{1}{8}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{15}$$

The operator $\hat{\mathbf{I}}_k^{k-1}$ is the straight injection.

The prolongation $\mathbf{I}_{k-1}^{k}$ is defined by bilinear interpolation in space. It is given in stencil form by

$$\mathbf{I}_{k-1}^{k} = \frac{1}{4}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \tag{16}$$

If $s=1$ no interpolation in time is needed, whereas if $s=2$ then $\mathbf{I}_{k-1}^{k}$ represents bilinear interpolation also in time.

Notice that the convergence behavior of the multigrid schemes presented in this paper remains qualitatively similar for various choices of prolongation and restriction operators. In particular one can use the upwinded versions of (15) and (16) defined in [13].

The choice of the smoothing operator $\mathbf{S}_k$ is a delicate one. We need guarantee a robust coupling between state and control variables and take care of the fact that the state variable evolves forward in time while the adjoint variable evolves backwards. One could first compute the state variable for all time steps with an initial control function and subsequently perform backward evolution of the adjoint variable and update the control. An iteration step of this type has been considered in [8]. This approach requires existence of global solutions in time of the uncontrolled problem and therefore cannot be applied for solving singular optimal control problems with possible finite-time blow-up [16]. For the same reason the method given in [9] cannot be applied.

In the following we describe our smoothing iteration which does not present the above-mentioned limitation and results advantageous in terms of robustness and efficiency also when solving linear problems.

Consider a collective Gauss–Seidel iteration which is applied at each grid point to the set of variables $\boldsymbol{\phi}_k = (y_k, p_k)$. For this purpose denote with $\mathbf{E}(\boldsymbol{\phi}_{ijm}) = [\mathbf{f} - \mathbf{A}(\boldsymbol{\phi})]_{ijm} = 0$, the two algebraic equations (10) and (11) for the two variables y_{ijm} and p_{ijm} at the grid point ijm. Further denote with $\mathbf{E}'$ the Jacobian of $\mathbf{E}$ with respect to these two variables. A sweep of the collective Gauss–Seidel

scheme is given by

$$\boldsymbol{\phi}_{ijm}^{(1)} = \boldsymbol{\phi}_{ijm}^{(0)} - [\mathbf{E}'(\boldsymbol{\phi}_{ijm}^{(0)})]^{-1}\mathbf{E}(\boldsymbol{\phi}_{ijm}^{(0)}). \tag{17}$$

This iteration was successfully used in [3] to solve steady-state optimality systems relative to the optimal control of explosive phenomena.

In case of time-dependent phenomena, iteration (17) will eventually diverge because the information of the opposite orientation of the state equation and of the adjoint equation is not taken into account. To add this information we propose to use (17) to update the state component y marching in the forward direction and to update the adjoint variable p using (17) but marching backwards in time. In this way a robust iteration is obtained given by the following

Time-splitted collective Gauss——Seidel iteration (TS-CGS)

1. Set $\boldsymbol{\phi}^0 = \tilde{\boldsymbol{\phi}}$.
2. For $m = 2, \ldots, N_t$ do
3. For ij in lexicographic order do

$$y_{ijm}^{(1)} = y_{ijm}^{(0)} - [\mathbf{E}'(\boldsymbol{\phi}_{ijm})]^{-1}\mathbf{E}(\boldsymbol{\phi}_{ijm})|_y,$$

$$p_{ijN_t-m+2}^{(1)} = p_{ijN_t-m+2}^{(0)} - [\mathbf{E}'(\boldsymbol{\phi}_{ijN_t-m})]^{-1}\mathbf{E}(\boldsymbol{\phi}_{ijN_t-m+2})|_p,$$

4. end.

Here, $[\mathbf{E}'(\boldsymbol{\phi}_{ijm'})]^{-1}\mathbf{E}(\boldsymbol{\phi}_{ijm'})$ is a two-component column vector corresponding to the variables y and p. A more explicit form of Step 3 follows

$$y_{ijm}^{(1)} = y_{ijm}^{(0)} - \left.\frac{(1+4\gamma)r_y(\boldsymbol{\phi}) - (\delta t/\nu)r_p(\boldsymbol{\phi})}{(1+4\gamma)^2 + (\delta t^2/\nu)}\right|_{ijm}^{(0)},$$

$$p_{ijN_t-m+2}^{(1)} = p_{ijN_t-m+2}^{(0)} - \left.\frac{(1+4\gamma)r_p(\boldsymbol{\phi}) + \delta t r_y(\boldsymbol{\phi})}{(1+4\gamma)^2 + (\delta t^2/\nu)}\right|_{ijN_t-m+2}^{(0)},$$

where $r_y(\boldsymbol{\phi})$ denotes the residual of (10) and $r_p(\boldsymbol{\phi})$ denotes the residual of (11) prior update. Obvious modifications are required to define time-splitted Red-Black collective Gauss–Seidel scheme or time-splitted collective Jacobi scheme.

The multigrid methods discussed in this section are designed to solve parabolic optimal control problems where the time discretization is by backward Euler scheme. In case of Crank–Nicolson discretization our multigrid approach can be successfully applied only for a small range of values of $\gamma \approx 1$. For $\gamma \gg 1$ our multigrid schemes possibly diverge. This fact is in agreement with results in [21] where it is shown that for large values of γ space–time multigrid solvers of parabolic problems are not robust in solving for Crank–Nicolson discretization. Following [21] three-level backward Euler discretization could be used for large γ when second-order time accuracy is required.

4. Numerical experiments

Purpose of the following experiments is to investigate the convergence properties of the multigrid methods described above. We use the FAS $V(1,1)$-cycle with the time-splitted collective

Table 2
Results of experiments with semicoarsening

$N_x \times N_y \times N_t$	γ	ρ	$\|y - \tilde{z}\|_0$	$\|r_y(\phi)\|_0$, $\|r_p(\phi)\|_0$
$v = 10^{-4}$				
$32 \times 32 \times 64$	16	0.146	$1.55 \cdot 10^{-3}$	$4.5 \cdot 10^{-10}$, $7.6 \cdot 10^{-12}$
$64 \times 64 \times 64$	64	0.164	$1.55 \cdot 10^{-3}$	$9.1 \cdot 10^{-10}$, $1.0 \cdot 10^{-11}$
$128 \times 128 \times 64$	256	0.159	$1.55 \cdot 10^{-3}$	$1.1 \cdot 10^{-9}$, $8.1 \cdot 10^{-12}$
$v = 10^{-6}$				
$32 \times 32 \times 64$	16	0.147	$4.03 \cdot 10^{-5}$	$1.4 \cdot 10^{-10}$, $1.9 \cdot 10^{-13}$
$64 \times 64 \times 64$	64	0.140	$4.23 \cdot 10^{-5}$	$2.6 \cdot 10^{-10}$, $2.1 \cdot 10^{-13}$
$128 \times 128 \times 64$	256	0.165	$4.27 \cdot 10^{-5}$	$3.3 \cdot 10^{-10}$, $5.8 \cdot 10^{-13}$
$v = 10^{-8}$				
$32 \times 32 \times 64$	16	0.008	$9.09 \cdot 10^{-7}$	$4.7 \cdot 10^{-15}$, $1.1 \cdot 10^{-18}$
$64 \times 64 \times 64$	64	0.06	$1.73 \cdot 10^{-6}$	$9.1 \cdot 10^{-12}$, $7.6 \cdot 10^{-16}$
$128 \times 128 \times 64$	256	0.134	$2.06 \cdot 10^{-6}$	$9.1 \cdot 10^{-11}$, $8.1 \cdot 10^{-15}$

Gauss-Seidel scheme as the smoothing iteration. The discretization parameters are chosen such that $\gamma = \delta t/h^2 \gg 1$, which is the most common situation where implicit time discretization is chosen.

To describe the results of the experiments we need to define some quantities. The multigrid convergence factor ρ is defined as the "asymptotic" value of the ratio of the norm of the dynamic residuals given by $\|r_y(\phi)\|_0 + \|r_p(\phi)\|_0/v$ resulting from two successive multigrid cycles. The tracking ability of our algorithms will be expressed in terms of the norm of the tracking error, $\|y - \tilde{z}\|_0$. Results reported in the tables are obtained with $N = 10$ FAS $V(1,1)$ cycles.

Let us consider $\Omega = (0,1) \times (0,1)$, $T = 1$, and the following objective function

$$z(\mathbf{x}, t) = (x_1 - x_1^2)(x_2 - x_2^2)\cos(4\pi t).$$

We take $y_0(\mathbf{x}) = z(\mathbf{x}, 0)$.

In Table 2 we report results of numerical experiments with multigrid using semicoarsening strategy. Three different grids with increasing refinement in space directions are considered. Similar results are obtained also on meshes with $N_t = \{128, 256, 512\}$. The observed convergence factors demonstrate usual multigrid convergence speeds and appear to be almost independent of the value of γ and weakly dependent on the value of v. This fact shows robustness of the multigrid scheme with TS-CGS smoothing. As the value of v increases then larger values of $\|y - \tilde{z}\|_0$ are obtained.

In case of multigrid coarsening in space and in time we observe slow convergence; see Table 3. Nevertheless, for sufficiently small values of v typical multigrid convergence factors are obtained. This fact is confirmed by the Fourier analysis presented later in this paper.

The setting $s = 2$ provides acceptable multigrid convergence rates also in case where T and correspondingly γ tend to be small. Small time intervals are of interest when considering control of transient phenomena. For example, consider the case $T = 0.01$ and a highly oscillating objective function given by

$$\tilde{z}(\mathbf{x}, t) = (x_1 - x_1^2)(x_2 - x_2^2)\sin(100\pi t).$$

Table 3
Results of experiments with standard coarsening

$N_x \times N_y \times N_t$	γ	ρ	$\|y - \tilde{z}\|_0$	$\|r_y(\phi)\|_0$, $\|r_p(\phi)\|_0$
$v = 10^{-4}$				
$32 \times 32 \times 64$	16	0.929	$1.60 \cdot 10^{-3}$	$1.9 \cdot 10^{-2}$, $2.6 \cdot 10^{-4}$
$64 \times 64 \times 64$	64	0.976	$1.69 \cdot 10^{-3}$	$7.3 \cdot 10^{-2}$, $5.5 \cdot 10^{-4}$
$128 \times 128 \times 64$	256	0.999	$2.20 \cdot 10^{-3}$	$1.4 \cdot 10^{-1}$, $1.2 \cdot 10^{-3}$
$v = 10^{-6}$				
$32 \times 32 \times 64$	16	0.905	$4.07 \cdot 10^{-5}$	$3.1 \cdot 10^{-3}$, $1.9 \cdot 10^{-6}$
$64 \times 64 \times 64$	64	0.140	$4.23 \cdot 10^{-5}$	$2.6 \cdot 10^{-10}$, $2.1 \cdot 10^{-13}$
$128 \times 128 \times 64$	256	0.165	$4.27 \cdot 10^{-5}$	$3.3 \cdot 10^{-10}$, $5.8 \cdot 10^{-13}$
$v = 10^{-8}$				
$32 \times 32 \times 64$	16	0.008	$9.09 \cdot 10^{-7}$	$4.7 \cdot 10^{-15}$, $1.1 \cdot 10^{-18}$
$64 \times 64 \times 64$	64	0.06	$1.73 \cdot 10^{-6}$	$9.1 \cdot 10^{-12}$, $7.6 \cdot 10^{-16}$
$128 \times 128 \times 64$	256	0.134	$2.06 \cdot 10^{-6}$	$9.1 \cdot 10^{-11}$, $8.1 \cdot 10^{-15}$

Table 4
Results of experiments with standard time coarsening for the case $T = 0.01$ and $\tilde{z}$ as objective function

v	ρ	$\|y - \tilde{z}\|_0$	$\|r_y(\phi)\|_0$, $\|r_p(\phi)\|_0$
10^{-3}	0.154	$2.29 \cdot 10^{-3}$	$1.1 \cdot 10^{-8}$, $3.2 \cdot 10^{-10}$
10^{-5}	0.117	$9.63 \cdot 10^{-4}$	$2.2 \cdot 10^{-9}$, $6.0 \cdot 10^{-12}$
10^{-7}	0.110	$4.61 \cdot 10^{-5}$	$1.4 \cdot 10^{-9}$, $3.7 \cdot 10^{-13}$

The finest grid is $N_x \times N_y \times N_t = 64 \times 64 \times 128$.

In this case the choices $s = 2$ gives good results even for moderate values of v as can be seen in Table 4.

Further numerical experiments demonstrate that the multigrid convergence behavior as observed in this section appears to be insensitive to the particular choice of the objective function, which may not be attainable.

5. Fourier mode analysis of the twogrid method

In this section, we use Fourier mode analysis [6,13,19,21] to analyze the convergence properties of the twogrid version of our parabolic optimal control solvers. Our aim is to investigate the dependence of the twogrid convergence factor on the weight v and on the ratio $\gamma = \delta t/h^2$. The Fourier analysis provides convergence results that closely predict the convergence factors observed experimentally. Notice that one space dimension is considered.

We use Fourier mode analysis assuming infinite grids. On the fine grid consider the Fourier components $\phi(\boldsymbol{j},\boldsymbol{\theta}) = \mathrm{e}^{\mathrm{i}\boldsymbol{j}\cdot\boldsymbol{\theta}}$ where i is the imaginary unit, $\boldsymbol{j} = (j_x, j_t) \in \mathbb{Z}\times\mathbb{Z}$, $\boldsymbol{\theta} = (\theta_x, \theta_t) \in [-\pi,\pi)^2$, and $\boldsymbol{j}\cdot\boldsymbol{\theta} = j_x\theta_x + j_t\theta_t$.

First, consider the case of semicoarsening in space. The frequency domain is spanned by the following two sets of frequencies

$$\boldsymbol{\theta}^{(0,0)} := (\theta_x, \theta_t) \quad \text{and} \quad \boldsymbol{\theta}^{(1,0)} := (\bar{\theta}_x, \theta_t),$$

where $(\theta_x,\theta_t) \in ([-\pi/2,\pi/2)\times[-\pi,\pi))$ and $\bar{\theta}_x = \theta_x - sign(\theta_x)\pi$. The components $\phi(\cdot,\boldsymbol{\theta}^{\boldsymbol{\alpha}})$ are called harmonics. The first harmonics $\phi(\cdot,\boldsymbol{\theta}^{(0,0)})$ represents low frequencies components in space. The second harmonics $\phi(\cdot,\boldsymbol{\theta}^{(1,0)})$ contains the high frequencies components in space direction. Both have all frequencies components in time direction. Using semicoarsening, we have that $\phi(\boldsymbol{j},\boldsymbol{\theta}^{(0,0)}) = \phi(\boldsymbol{j},\boldsymbol{\theta}^{(1,0)})$ on the coarse grid. The purpose of multigrid is to reduce the high frequency error components by applying the smoothing operator $\mathbf{S}_k$ and to reduce the low frequency error components by coarse grid correction given by

$$\boldsymbol{CG}_k^{k-1} = [\mathbf{I}_k - \mathbf{I}_{k-1}^k(\mathbf{A}_{k-1})^{-1}\mathbf{I}_k^{k-1}\mathbf{A}_k].$$

Denote with $E_k^\theta = \mathrm{span}[\phi_k(\cdot,\boldsymbol{\theta}^{\boldsymbol{\alpha}}) : \boldsymbol{\alpha} \in \{(0,0),(1,0)\}]$. In the linear case, under the assumption that all multigrid components are linear and that $(\mathbf{A}_{k-1})^{-1}$ exists, we have a representation of the twogrid operator $\mathbf{TG}_k^{k-1}$ on the space $E_k^\theta \times E_k^\theta$ by a 4×4 matrix given by

$$\widehat{\boldsymbol{TG}}_k^{k-1}(\boldsymbol{\theta}) = \hat{\mathbf{S}}_k(\boldsymbol{\theta})^{m_2}\widehat{\boldsymbol{CG}}_k^{k-1}(\boldsymbol{\theta})\hat{\mathbf{S}}_k(\boldsymbol{\theta})^{m_1},$$

where the hat denotes the Fourier symbol [19] of the given operator.

To determine the explicit form of the operator symbols given above, consider the action of the operators on the couple $(\tilde{y},\tilde{p}) \in E_k^\theta \times E_k^\theta$, where:

$$\tilde{y} = \sum_{p=0,1} Y^{(p,0)}\phi_k(\boldsymbol{j},\boldsymbol{\theta}^{(p,0)}) \quad \text{and} \quad \tilde{p} = \sum_{p=0,1} P^{(p,0)}\phi_k(\boldsymbol{j},\boldsymbol{\theta}^{(p,0)}). \tag{18}$$

Here, $Y^{\boldsymbol{\alpha}}$ and $P^{\boldsymbol{\alpha}}$ are the Fourier amplitudes.

Consider to apply the TS-CGS step first to all state variables leaving the adjoint variables unchanged and then vice versa. Under this assumption substituting (18) into (10) and (11) we obtain

$$\hat{\mathbf{S}}_k(\boldsymbol{\theta}) = \mathrm{diag}\{\sigma(\boldsymbol{\theta}^{(0,0)}),\sigma(\boldsymbol{\theta}^{(1,0)}),\sigma(\boldsymbol{\theta}^{(0,0)}),\sigma(\boldsymbol{\theta}^{(1,0)})\},$$

where

$$\sigma(\boldsymbol{\theta}^{(p,q)}) = \frac{\nu\gamma(2\gamma+1)\mathrm{e}^{\mathrm{i}\theta_x^p}}{\delta t^2 + \nu[(2\gamma+1)^2 - \gamma(2\gamma+1)\mathrm{e}^{-\mathrm{i}\theta_x^p} - (2\gamma+1)\mathrm{e}^{-\mathrm{i}\theta_t^q}]}.$$

In the following we choose $\delta t = 1/64$ corresponding to the experimental setting.

A way to characterize the smoothing property of the operator $\mathbf{S}_k$ is to assume an ideal coarse grid correction which annihilates the low frequency error components and leaves the high frequency error components unchanged. That is, one defines the projection operator Q_k^{k-1} on E_k^θ by

$$Q_k^{k-1}\phi(\boldsymbol{\theta},\cdot) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}^{(0,0)}, \\ \phi(\cdot,\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} = \boldsymbol{\theta}^{(1,0)}. \end{cases}$$

On the space $E_k^\theta \times E_k^\theta$ we then have

$$\hat{\boldsymbol{Q}}_k^{k-1}(\boldsymbol{\theta}) = \begin{bmatrix} Q_k^{k-1} & 0 \\ 0 & Q_k^{k-1} \end{bmatrix} \quad \text{for } \boldsymbol{\theta} \in ([-\pi/2, \pi/2) \times [-\pi, \pi)). \tag{19}$$

In this framework the smoothing property of $\mathbf{S}_k$ is defined as follows:

$$\mu = \max\{r(\hat{\boldsymbol{Q}}_k^{k-1}(\boldsymbol{\theta})\hat{\mathbf{S}}_k(\boldsymbol{\theta})):\ \boldsymbol{\theta} \in ([-\pi/2, \pi/2) \times [-\pi, \pi))\}, \tag{20}$$

where r is the spectral radius. Notice that, assuming an ideal coarse grid correction takes place, the convergence factor of the twogrid scheme is given by $\mu^{m_1+m_2}$.

In space, we consider a full-weighting restriction operator whose symbol is given by

$$\hat{\mathbf{I}}_k^{k-1}(\boldsymbol{\theta}) = \frac{1}{2}\begin{bmatrix} (1+\cos(\theta_x)) & (1-\cos(\theta_x)) & 0 & 0 \\ 0 & 0 & (1+\cos(\theta_x)) & (1-\cos(\theta_x)) \end{bmatrix}.$$

For the linear prolongation operator we have $\hat{\mathbf{I}}_{k-1}^k(\boldsymbol{\theta}) = \hat{\mathbf{I}}_k^{k-1}(\boldsymbol{\theta})^{\mathrm{T}}$. The symbol of the fine grid operator is

$$\hat{\mathbf{A}}_k(\boldsymbol{\theta}) = \begin{bmatrix} a_y(\boldsymbol{\theta}^{(0,0)}) & 0 & -\delta t/\nu & 0 \\ 0 & a_y(\boldsymbol{\theta}^{(1,0)}) & 0 & -\delta t/\nu \\ \delta t & 0 & a_p(\boldsymbol{\theta}^{(0,0)}) & 0 \\ 0 & \delta t & 0 & a_p(\boldsymbol{\theta}^{(1,0)}) \end{bmatrix},$$

where

$$a_y(\boldsymbol{\theta}^{(p,q)}) = 2\gamma\cos(\theta_x^p) - \mathrm{e}^{-\mathrm{i}\theta_t^q} - 2\gamma - 1 \quad \text{and} \quad a_p(\boldsymbol{\theta}^{(p,q)}) = 2\gamma\cos(\theta_x^p) - \mathrm{e}^{\mathrm{i}\theta_t^q} - 2\gamma - 1.$$

The symbol of the coarse grid operator follows:

$$\hat{\mathbf{A}}_{k-1}(\boldsymbol{\theta}) = \begin{bmatrix} \gamma\cos(2\theta_x)/2 - \mathrm{e}^{-\mathrm{i}\theta_t} - \gamma/2 - 1 & -\delta t/\nu \\ \delta t & \gamma\cos(2\theta_x)/2 - \mathrm{e}^{\mathrm{i}\theta_t} - \gamma/2 - 1 \end{bmatrix}.$$

Notice that on the coarser grid δt remains unchanged while $\gamma \to \gamma/4$ by coarsening.

Based on the representation on $\boldsymbol{TG}_k^{k-1}$ by a 4×4 matrix $\widehat{\boldsymbol{TG}}_k^{k-1}(\boldsymbol{\theta})$ we can calculate the convergence factor given by

$$\eta(\boldsymbol{TG}_k^{k-1}) = \sup\{r(\widehat{\boldsymbol{TG}}_k^{k-1}(\boldsymbol{\theta})):\ \boldsymbol{\theta} \in ([-\pi/2, \pi/2) \times [-\pi, \pi))\}.$$

By Fourier mode analysis the problem of determining the convergence factor of a twogrid scheme is reduced to that of determining the spectral radius of a 4×4 matrix. This task may be performed using any symbolic package (we use Mathematica). In Fig. 1 we report the smoothing factor and the convergence factor depending on the value of ν and on the value of γ for $\delta t = 1/64$. Similar figures are obtained with different choices of time-step size. We observe that the values of convergence factors predicted by Fourier mode analysis are very close to those obtained experimentally and reported in

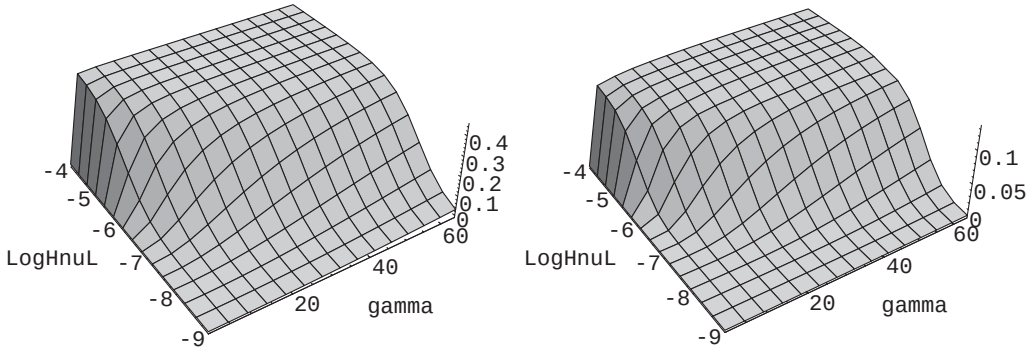

Fig. 1. Semicoarsening: (left) smoothing factor as a function of v and γ; (right) twogrid convergence factor as a function of v and γ ($m_1 = m_2 = 1$).

Table 5
η for semicoarsening; $\delta t = 1/64$

	v		
γ	10^{-4}	10^{-6}	10^{-8}
16	0.126	0.102	$4.28 \cdot 10^{-4}$
32	0.130	0.120	$5.13 \cdot 10^{-3}$
48	0.131	0.127	$1.72 \cdot 10^{-2}$
64	0.132	0.130	$3.38 \cdot 10^{-2}$

Table 2. Convergence factor estimates are also reported in numerical form in Table 5 to facilitate comparison with the results of Table 2.

Next, we consider the case of standard coarsening. A four-dimensional space of harmonics has to be taken

$$E_k^\theta = \text{span}[\phi_k(\cdot, \boldsymbol{\theta}^{\boldsymbol{\alpha}}):\ \boldsymbol{\alpha} \in \{(0,0),(1,1),(1,0),(0,1)\}],$$

where for any low frequency $\boldsymbol{\theta} = (\theta_x, \theta_t) \in [-\pi/2, \pi/2)^2$, we have

$$\boldsymbol{\theta}^{(0,0)} := (\theta_x, \theta_t), \quad \boldsymbol{\theta}^{(1,1)} := (\bar{\theta}_x, \bar{\theta}_t),$$
$$\boldsymbol{\theta}^{(1,0)} := (\bar{\theta}_x, \theta_t), \quad \boldsymbol{\theta}^{(0,1)} := (\theta_x, \bar{\theta}_t).$$

Therefore the Fourier symbol of the smoothing operator is now a 8×8 diagonal matrix

$$\hat{\mathbf{S}}_k(\boldsymbol{\theta}) = \text{diag}\{\sigma(\boldsymbol{\theta}^{(0,0)}), \sigma(\boldsymbol{\theta}^{(1,1)}), \sigma(\boldsymbol{\theta}^{(1,0)}), \sigma(\boldsymbol{\theta}^{(0,1)}), \sigma(\boldsymbol{\theta}^{(0,0)}), \sigma(\boldsymbol{\theta}^{(1,1)}), \sigma(\boldsymbol{\theta}^{(1,0)}), \sigma(\boldsymbol{\theta}^{(0,1)})\}.$$

The operator $\hat{\mathbf{A}}_k(\boldsymbol{\theta})$ is given by

$$\begin{bmatrix} a_y(\boldsymbol{\theta}^{(0,0)}) & 0 & 0 & 0 & -\delta t/v & 0 & 0 & 0 \\ 0 & a_y(\boldsymbol{\theta}^{(1,1)}) & 0 & 0 & 0 & -\delta t/v & 0 & 0 \\ 0 & 0 & a_y(\boldsymbol{\theta}^{(1,0)}) & 0 & 0 & 0 & -\delta t/v & 0 \\ 0 & 0 & 0 & a_y(\boldsymbol{\theta}^{(0,1)}) & 0 & 0 & 0 & -\delta t/v \\ \delta t & 0 & 0 & 0 & a_p(\boldsymbol{\theta}^{(0,0)}) & 0 & 0 & 0 \\ 0 & \delta t & 0 & 0 & 0 & a_p(\boldsymbol{\theta}^{(1,1)}) & 0 & 0 \\ 0 & 0 & \delta t & 0 & 0 & 0 & a_p(\boldsymbol{\theta}^{(1,0)}) & 0 \\ 0 & 0 & 0 & \delta t & 0 & 0 & 0 & a_p(\boldsymbol{\theta}^{(0,1)}) \end{bmatrix}.$$

To write the symbol of the coarse grid operator notice that $\gamma \to \gamma/2$ and $\delta t \to 2\delta t$ by coarsening. So we have

$$\hat{\mathbf{A}}_{k-1}(\boldsymbol{\theta}) = \begin{bmatrix} \gamma\cos(2\theta_x) - \mathrm{e}^{-2\mathrm{i}\theta_t} - \gamma - 1 & -2\delta t/v \\ 2\delta t & \gamma\cos(2\theta_x) - \mathrm{e}^{2\mathrm{i}\theta_t} - \gamma - 1 \end{bmatrix}.$$

The symbol of the bi-linear interpolation operator is given by

$$\hat{\mathbf{I}}_{k-1}^{k}(\boldsymbol{\theta}) = \begin{bmatrix} I(\boldsymbol{\theta}^{(0,0)}) & I(\boldsymbol{\theta}^{(1,1)}) & I(\boldsymbol{\theta}^{(1,0)}) & I(\boldsymbol{\theta}^{(0,1)}) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I(\boldsymbol{\theta}^{(0,0)}) & I(\boldsymbol{\theta}^{(1,1)}) & I(\boldsymbol{\theta}^{(1,0)}) & I(\boldsymbol{\theta}^{(0,1)}) \end{bmatrix}^{\mathrm{T}},$$

where

$$I(\boldsymbol{\theta}^{\boldsymbol{\alpha}}) = \tfrac{1}{4}(1+\cos(\theta_x^{\alpha_x}))(1+\cos(\theta_t^{\alpha_t})).$$

For restriction we use full-weighting in the space direction, as in the semicoarsening case.

The smoothing factor of the smoothing iteration is again defined through the action of $\mathbf{S}_k$ on the high frequency components of the error as follows:

$$\mu = \max\{r(\hat{\boldsymbol{Q}}_k^{k-1}(\boldsymbol{\theta})\hat{\mathbf{S}}_k(\boldsymbol{\theta})):\ \boldsymbol{\theta} \in [-\pi/2, \pi/2)^2\}, \tag{21}$$

where $\hat{\boldsymbol{Q}}_k^{k-1}(\boldsymbol{\theta})$ represents the projection onto the space of high frequencies.

The convergence factor is given by

$$\eta(\boldsymbol{TG}_k^{k-1}) = \max\{r(\widehat{\boldsymbol{TG}}_k^{k-1}(\boldsymbol{\theta})):\ \boldsymbol{\theta} \in [-\pi/2, \pi/2)^2\}.$$

The values of the smoothing factor and of the convergence factor as functions of v and γ are reported in Fig. 2 (for $\delta t = 1/64$); see also Table 6. Comparing with the semicoarsening case the smoothing factor and correspondingly the convergence factor have deteriorated. Nevertheless, as γ (correspondingly δt) and v are sufficiently small these factors improve. Also in the standard coarsening case Fourier mode analysis is able to predict correctly the convergence behavior of the multigrid scheme; compare Tables 3 and 4 with Fig. 2.

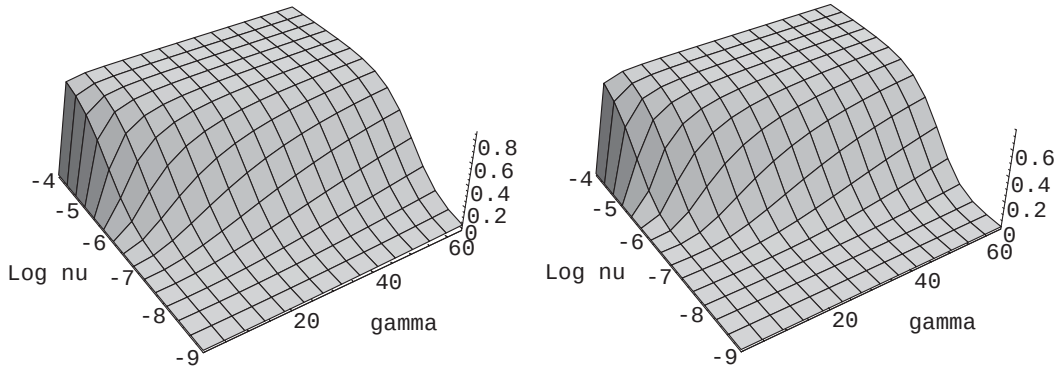

Fig. 2. Standard coarsening: (left) smoothing factor as a function of ν and γ; (right) twogrid convergence factor as a function of ν and γ ($m_1 = m_2 = 1$).

Table 6
η for standard coarsening, $\delta t = 1/64$

	ν		
γ	10^{-4}	10^{-6}	10^{-8}
16	0.731	0.386	$4.46 \cdot 10^{-4}$
32	0.733	0.613	$6.06 \cdot 10^{-3}$
48	0.734	0.676	$2.48 \cdot 10^{-2}$
64	0.734	0.700	$6.03 \cdot 10^{-2}$

6. Constrained optimal control problems

While we consider mainly the case of unconstrained control, results of experiments with constrained control problems are also reported. In [2] two multigrid approaches were investigated. The first one considers the presence of constraints as a generic nonlinearity within the FAS multigrid approach [6]. The second one uses the primal-dual strategy [1] and defines a primal-dual-multigrid algorithm. In this section we report further results regarding the former 'direct' multigrid approach which appears to be computationally more convenient than the primal-dual-multigrid strategy.

Consider the optimal control problem (1) with the control u belonging to the following admissible set

$$u \in \{v \in L^2(Q) \,|\, c_1 \leqslant v(\mathbf{x}) \leqslant c_2 \text{ a.e. in } Q\}, \tag{22}$$

Existence of a unique solution can be proved [15] and is characterized by the following optimality system

$$-\partial_t y + \Delta y = u, \tag{23}$$

$$\partial_t p + \Delta p + (y - z) = 0, \tag{24}$$

$$u - \sup\left(c_1, \inf\left(c_2, \frac{p}{\nu}\right)\right) = 0. \tag{25}$$

Table 7
Results of experiments with semicoarsening. $V(2,2)$-cycle; $c_1 = -10$ and $c_2 = 10$

$N_x \times N_y \times N_t$	γ	ρ	$\|y - \tilde{z}\|_0$	$\|r_y(\phi)\|_0$, $\|r_p(\phi)\|_0$
$v = 10^{-4}$				
$64 \times 64 \times 64$	64	0.133	$6.00 \cdot 10^{-2}$	$5.3 \cdot 10^{-9}$, $2.4 \cdot 10^{-11}$
$128 \times 128 \times 128$	128	0.137	$5.90 \cdot 10^{-2}$	$1.0 \cdot 10^{-8}$, $5.6 \cdot 10^{-11}$
$v = 10^{-6}$				
$64 \times 64 \times 64$	64	0.770	$5.45 \cdot 10^{-2}$	$7.9 \cdot 10^{-2}$, $6.7 \cdot 10^{-4}$
$128 \times 128 \times 128$	128	0.761	$5.37 \cdot 10^{-2}$	$6.6 \cdot 10^{-2}$, $6.0 \cdot 10^{-4}$

The discretization scheme (10) and (11) applies also to (23), (24), and (25) with the right-hand side of (10) being replaced by $\delta t G(p_{ijm})$ where

$$G(p_{ijm}) = \max\left(c_1, \min\left(c_2, \frac{p_{ijm}}{v}\right)\right). \tag{26}$$

Unfortunately the presence of the term (26) prevents us from defining $\mathbf{E}'$ in the obvious way. Specifically, we cannot differentiate (25) with respect to p. To overcome this difficulty we set $G'(p) = 1/v$ as in the unconstrained case in those part of the domain where the constrains are inactive. Whenever a constraint is active we set $G'(p) = 0$. So the TS-CGS scheme is as follows:

$$y^{(1)}_{ijm} = y^{(0)}_{ijm} - \frac{(1+4\gamma) r_y(\phi) - \delta t G'(p) r_p(\phi)}{(1+4\gamma)^2 + \delta t^2 G'(p)}\bigg|^{(0)}_{ijm},$$

$$p^{(1)}_{ijN_t-m+2} = p^{(0)}_{ijN_t-m+2} - \frac{(1+4\gamma) r_p(\phi) + \delta t r_y(\phi)}{(1+4\gamma)^2 + \delta t^2 G'(p)}\bigg|^{(0)}_{ijN_t-m+2},$$

where $m = 2, \ldots, N_t$.

Using this relaxation scheme a lack of robustness of the V-cycle with respect to changes of the value of the weight of the cost of the control may be observed [2]. Specifically, as v tends to be smaller the convergence factor of the V-cycle version of our multigrid scheme worsen. We now show that much better computational properties are obtained when using the W-cycle version of the our algorithms.

Consider the following objective function

$$z(\mathbf{x}, t) = \sin(2\pi t)\sin(\pi x_1)\sin(\pi x_2).$$

For the purpose of comparison we report in Table 7 results of experiments using the V-cycle algorithm with semicoarsening. Observe the deterioration of the convergence factor when reducing the value of v. A different behavior is observed using the W-cycle scheme. For this latter case results are reported in Table 8. In the W-cycle case much better convergence factors can be observed and the algorithm remains competitive also for small values of v. The use of W-cycles is also beneficial for standard coarsening multigrid schemes. Nevertheless the semicoarsening-based algorithm outperforms the standard-coarsening scheme.

The time evolution for the present constrained optimal control problem is represented in Fig. 3. In this figure the values of the state variable, of the objective function, and of the resulting optimal

Table 8
Results of experiments with semicoarsening. $W(2,2)$-cycle; $c_1 = -10$ and $c_2 = 10$

$N_x \times N_y \times N_t$	γ	ρ	$\|y - \tilde{z}\|_0$	$\|r_y(\phi)\|_0$, $\|r_p(\phi)\|_0$
$v = 10^{-4}$				
$64 \times 64 \times 64$	64	0.056	$6.00 \cdot 10^{-2}$	$2.3 \cdot 10^{-13}$, $1.8 \cdot 10^{-15}$
$128 \times 128 \times 128$	128	0.052	$5.90 \cdot 10^{-2}$	$9.0 \cdot 10^{-13}$, $7.2 \cdot 10^{-15}$
$v = 10^{-6}$				
$64 \times 64 \times 64$	64	0.393	$5.61 \cdot 10^{-2}$	$1.2 \cdot 10^{-4}$, $7.0 \cdot 10^{-7}$
$128 \times 128 \times 128$	128	0.143	$5.50 \cdot 10^{-2}$	$1.4 \cdot 10^{-8}$, $2.7 \cdot 10^{-11}$

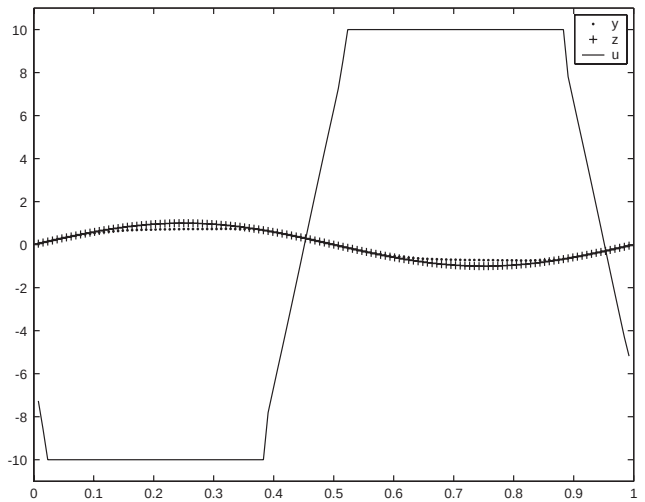

Fig. 3. Solution of the constrained optimal control problem ($v = 10^{-6}$). Dotted line: $y(0.5, 0.5, t)$; '+' line: $z(0.5, 0.5, t)$; solid line: $u(0.5, 0.5, t)$.

control at $(x_1, x_2) = (0.5, 0.5)$ for all time steps are depicted. Observe the ability of the multigrid solution to track the given objective function.

7. Conclusions

In this paper, the solution of parabolic distributed optimal control problems discretized by finite differences and solved by means of space–time multigrid techniques was presented. Accuracy properties of finite difference approximation were discussed and validated numerically. The multigrid schemes presented in this paper were based on a robust smoothing iteration and two different coarsening strategies: semicoarsening and standard coarsening. These algorithms showed robustness with respect to changes of the value of the weight of the cost of the control. In case of semicoarsening typical multigrid convergence factors were obtained in all cases. The Fourier mode analysis presented

in this paper provided justification of the multigrid convergence behavior observed with numerical experiments. It was shown that the use of W-cycle multigrid schemes is advantageous when solving constrained parabolic optimal control problems.

References

[1] M. Bergounioux, K. Ito, K. Kunisch, Primal-dual strategy for constrained optimal control problems, SIAM J. Control Optim. 37 (4) (1999) 1176–1194.
[2] A. Borzì, Fast multigrid methods for parabolic optimal control problems, Proceedings of the 18th GAMM-Seminar, Leipzig, 2002, pp. 1–10.
[3] A. Borzì, K. Kunisch, The numerical solution of the steady state solid fuel ignition model and its optimal control, SIAM J. Sci. Comput. 22 (1) (2000) 263–284.
[4] A. Borzì, K. Kunisch, A multigrid method for optimal control of time-dependent reaction diffusion processes, in: K.H. Hoffmann, R. Hoppe, V. Schulz (Eds.), Fast Solution of Discretized Optimization Problems, International Series on Numerical Mathematics, Vol. 138, Birkhäuser, Basel, 2001.
[5] A. Borzì, K. Kunisch, M. Vanmaele, A multi-grid approach to the optimal control of solid fuel ignition problems, in: E. Dick, K. Riemslagh, J. Vierendeels (Eds.), Lecture Notes in Computer Science and Engineering, European Multigrid Meeting, 1999, Springer, Berlin, 2000.
[6] A. Brandt, Multi-level adaptive solutions to boundary-value problems, Math. Comp. 31 (1977) 333–390.
[7] A. Brandt, J. Greenwald, Parabolic multigrid revisited, in: W. Hackbusch, U. Trottenberg (Eds.), Multigrid Methods, Vol. III, International Series of Numerical Mathematics, Vol. 98, Birkhäuser, Basel, 1991.
[8] W. Hackbusch, A numerical method for solving parabolic equations with opposite orientations, Computing 20 (1978) 229–240.
[9] W. Hackbusch, Numerical solution of linear and nonlinear parabolic optimal control problems, in: Lecture Notes in Control and Information Science, Vol. 30, Springer, Berlin, 1981.
[10] W. Hackbusch, Parabolic multigrid methods, in: R. Glowinski, J.-L. Lions (Eds.), Computing Methods in Applied Sciences and Engineering, Vol. VI, North-Holland, Amsterdam, 1984.
[11] W. Hackbusch, Multi-Grid Methods and Applications, Springer, New York, 1985.
[12] W. Hackbusch, Elliptic Differential Equations, Springer, New York, 1992.
[13] G. Horton, S. Vandewalle, A space–time multigrid method for parabolic partial differential equations, SIAM J. Sci. Comput. 16 (4) (1995) 848–864.
[14] K. Ito, K. Kunisch, Asymptotic properties of receding horizon optimal control problems, SIAM J. Control Optim. 40 (5) (2002) 1585–1610.
[15] J.L. Lions, Optimal Control of Systems Governed by Partial Differential Equations, Springer, Berlin, 1971.
[16] J.L. Lions, Control of Distributed Singular Systems, Gauthier-Villars, Paris, 1985.
[17] K. Malanowski, Convergence of approximations vs. regularity of solutions for convex, control-constrained optimal-control problems, Appl. Math. Optim. 8 (1981) 69–95.
[18] P. Neittaanmäki, D. Tiba, Optimal Control of Nonlinear Parabolic Systems, Marcel Dekker, New York, 1994.
[19] U. Trottenberg, C. Oosterlee, A. Schüller, Multigrid, Academic Press, London, 2001.
[20] S. Vandewalle, Parallel Multigrid Waveform Relaxation for Parabolic Problems, B.G. Teubner, Stuttgart, 1993.
[21] S. Vandewalle, G. Horton, Fourier mode analysis of the multigrid waveform relaxation and time-parallel multigrid methods, Computing 54 (4) (1995) 317–330.
[22] S. Vandewalle, R. Piessens, Efficient parallel algorithms for solving initial-boundary value and time-periodic parabolic partial differential equations, SIAM J. Sci. Statist. Comput. 13 (6) (1992) 1330–1346.